

On Bayesian Nonparametric Continuous Time Series Models

George Karabatsos¹

University of Illinois-Chicago

and

Stephen G. Walker²

University of Kent, United Kingdom

March 2, 2013

Abstract: This paper is a note on the use of Bayesian nonparametric mixture models for continuous time series. We identify a key requirement for such models, and then establish that there is a single type of model which meets this requirement. As it turns out, the model is well known in multiple change-point problems.

Keywords: Change point; Mixture model.

¹Corresponding author. Professor, University of Illinois-Chicago, U.S.A., Program in Measurement and Statistics. 1040 W. Harrison St. (MC 147), Chicago, IL 60607. E-mail: gkarabatsos1@gmail.com. Phone: 312-413-1816.

²Professor, University of Kent, United Kingdom, School of Mathematics, Statistics & Actuarial Science. Currently, Visiting Professor at The University of Texas at Austin, Division of Statistics and Scientific Computation.

1 Introduction

A lot of recent research has focused on the development of Bayesian nonparametric, countably-infinite mixture models for time series data. This work has aimed to relax the normality assumptions of the general class of dynamic linear models (West & Harrison, 1997), which already encompasses traditional normal (time-static) linear regression, autoregressions, autoregressive moving average (ARMA) models, and nonstationary polynomial trend and time-series models.

These Bayesian nonparametric (infinite-mixture) time series models have the general form:

$$f(y_t|t) = \int f(y_t|\mathbf{x}_t, \boldsymbol{\gamma}, \boldsymbol{\theta}) dG_t(\boldsymbol{\theta}) = \sum_{j=1}^{\infty} f(y_t|\mathbf{x}_t, \boldsymbol{\gamma}, \boldsymbol{\theta}_{tj}) \omega_j(t),$$

given time $t \in \mathcal{T}$; kernel (component) densities $\{f(\cdot|\mathbf{x}_t, \boldsymbol{\gamma}, \boldsymbol{\theta}_{tj})\}_{j=1}^{\infty}$ which are often specified by normal densities of a dynamic linear model; mixture distribution

$$G_t(\cdot) = \sum_{j=1}^{\infty} \omega_j(t) \delta_{\boldsymbol{\theta}_{tj}}(\cdot),$$

which is formed by an infinite-mixture of point-mass distributions $\delta_{\boldsymbol{\theta}_{tj}}(\cdot)$ with mixture weights $\omega_j(t)$; and prior distributions $\boldsymbol{\gamma} \sim \pi(\boldsymbol{\gamma})$, $\boldsymbol{\theta}_{tj} \sim G_{0t}$, $\{\omega_j(t)\}_{j=1}^{\infty} \sim \Pi$. All of the earlier models focuses on discrete time, and specify G_t to be some variant of the Dependent Dirichlet process (DDP) (MacEachern, 1999, 2000, 2001), so that the mixture weights have a stick-breaking form, with

$$\omega_j(t) = v_j(t) \prod_{l=1}^{j-1} (1 - v_l(t)), \text{ and } v_j(t) : \mathcal{T} \rightarrow [0, 1].$$

(Sethuraman, 1994). Such DDP-based time-series models either assume time-dependent stick-breaking weights (Griffin & Steel, 2006, 2011; Rodriguez & Dunson, 2011), or assume non-time-dependent stick-breaking weights and a time-dependent prior (baseline) distribution G_{0t} (Rodriguez & ter Horst, 2008), or assume a fully non-time-dependent Dirichlet process (DP) $G_t = G$ with only time-dependence in the kernel densities (Hatjispyros, et al., 2009; Tang & Ghosal, 2007; Lau & So, 2008; Caron et al., 2008; Giardina et al., 2011; Di Lucca et al., 2012). Other related approaches construct a time-dependent DDP G_t either by generalizing the Pólya urn scheme of the DP (e.g., Zhu et al., 2005; Caron et al., 2007); by a convex combination of hierarchical Dirichlet processes (HDP) or DPs (Ren et al., 2008; Dunson, 2006); by a HDP-based hidden Markov model that has infinitely-many states (Fox et al., 2008, 2011); or by a Markov-switching model having finitely-many states (Taddy & Kottas, 2009).

The more recent work on Bayesian nonparametric time series modeling has focused on continuous time, and on developing a time-dependent mixture distribution that has the general form,

$$G_t = \sum_{j=1}^{\infty} \omega_j(t) \delta_{\boldsymbol{\theta}_j}(\cdot), \tag{1}$$

based on a process other than the DDP. Above, a baseline prior $\boldsymbol{\theta}_j \sim_{iid} G_0$ is assumed, which is a standard assumption.

In Section 2 we describe these continuous time series models, namely, the geometric model (Mena, Ruggiero, & Walker, 2011), and a normalized random measure model (NRM) (Griffin, 2011). In Section 3, we highlight a key property such models are required to possess, and we identify a necessary model which has such a property. We also in this section prove that the existing continuous time series models do not have the required property.

2 Continuous time models

The geometric model constructs a dependent process G_t using time-dependent geometric mixture weights

$$\omega_j(t) = \lambda_t(1 - \lambda_t)^{j-1}, \quad (2)$$

with λ_t specified as a two-type Wright-Fisher diffusion (Mena, Ruggiero, & Walker, 2011).

The $(\lambda)_t$ follow a stochastic process with the stationary density being a $\text{beta}(a, b)$. The transition mechanism is given, for $t > s$, by

$$p(\lambda_t | \lambda_s) = \sum_{m=0}^{\infty} p_h(m) p(\lambda_t | m, \lambda_s)$$

where $h = t - s$ and

$$p(\lambda_t | m, \lambda_s) = \sum_{k=0}^m \text{beta}(\lambda_t | a + k, b + m - k) \text{bin}(k | m, \lambda_s)$$

and

$$p_h(m) = \frac{(a + b)_m \exp(-mch)}{m!} (1 - e^{-ch})^{a+b}$$

for some $c > 0$.

Hence G_t is a continuous time process and the properties are studied in Mena, Ruggiero, and Walker (2011).

The normalized random measures (NRM) model constructs a time-dependent process G_t using time-dependent mixture weights that are formed by normalizing a stochastic process derived from non-Gaussian Ornstein-Uhlenbeck processes (Griffin, 2011). Specifically, these weights are constructed by

$$\omega_j(t) = \frac{\mathbf{1}(\tau_j \leq t) \exp(-\lambda(t - \tau_j)) J_j}{\sum_{l=1}^{\infty} \mathbf{1}(\tau_l \leq t) \exp(-\lambda(t - \tau_l)) J_l}, \quad (3)$$

where (τ, J) follows a Poisson process with intensity $\lambda w(J)$, where w is a Lévy density.

Details and examples of obtaining the (τ_j, J_j) are provided by Griffin (2011). Aside from the specific examples considered in this paper, we also note that any sequence of (τ_j, J_j) are permissible provided

$$\sum_{l=1}^{\infty} \mathbf{1}(\tau_l \leq t) \exp(-\lambda(t - \tau_l)) J_l < \infty$$

for all t .

3 A key property

Using the mixture model

$$f(y|t) = \int K(y|\theta) G_t(d\theta) = \sum_{j=1}^{\infty} w_j(t) K(y|\theta_j),$$

we insist on the obvious requirement that for all suitably small h , we want y_t and y_{t+h} to be arising from the same component. This requirement is clearly not met by simply insisting that $G_{t+h} \rightarrow G_t$ as $h \rightarrow 0$.

So, in this paper, we introduce the argument that a Bayesian nonparametric continuous time series model should have a certain property. Specifically, based on the above discussion, we need the property that

$$P(\theta_t = \theta_{t+h}) \rightarrow 1 \quad \text{as } h \rightarrow 0,$$

where θ_t denotes a sample from

$$G_t = \sum_{j=1}^{\infty} \omega_j(t) \delta_{\theta_j},$$

i.e. that $\theta_t|G_t \sim G_t$, which means that $P(\theta_t = \theta_j) = w_j(t)$.

Now it can be shown that

$$P(\theta_t = \theta_{t+h}) = \sum_{j=1}^{\infty} P(\theta_t = \theta_{t+h} = \theta_j),$$

and hence we are asking for

$$E \left\{ \sum_{j=1}^{\infty} \omega_j(t) \omega_j(t+h) \right\} \rightarrow 1 \quad \text{as } h \rightarrow 0.$$

For this, it is necessary that

$$D(h) = \sum_{j=1}^{\infty} \omega_j(t) \omega_j(t+h) \rightarrow 1 \quad \text{in probability as } h \rightarrow 0.$$

Now assume that

$$\sup_j |\omega_j(t+h) - \omega_j(t)| \rightarrow 0 \quad \text{a.s. as } h \rightarrow 0$$

which is an extremely mild condition.

Hence, for any $\epsilon > 0$,

$$\sup_j |\omega_j(t+h) - \omega_j(t)| < \epsilon$$

for all small enough h . Therefore, for all small h , we have

$$D(h) \leq \sum_{j=1}^{\infty} \omega_j^2(t) + \epsilon \quad \text{a.s.}$$

The only way we can now recover the convergence to 1 in probability is that

$$\omega_j(t) = 1 \quad \text{a.s.}$$

for a particular j , which will depend on t .

Hence, we believe that a Bayesian nonparametric continuous time series model should specify a time-dependent mixture distribution G_t of the type given in (1), where

$$\omega_j(t) = \mathbf{1}(t \in A_j),$$

and the $(A_j)_j$ form a random partition of $(0, \infty)$. In other words, we recommend Bayesian nonparametric change-point mode for time series analysis. Specifically, let $\mathcal{D} = \{(y_{t_i})\}_{i=1}^n$ denote a sample of data consisting of n dependent responses y_{t_i} observed at time points t_i . Then, such a model may be specified as:

$$y_{t_i} \sim f(y_{t_i} | \boldsymbol{\theta}_{z(t_i)}), \quad i = 1, \dots, n, \quad (4a)$$

$$z(t_i) = j \iff \tau_{j-1} < t_i \leq \tau_j = (\tau_{j-1} + \epsilon_j) \quad (4b)$$

$$\epsilon_j \sim \text{Ex}(\lambda), \quad j = 1, 2, \dots, \quad (4c)$$

$$\boldsymbol{\theta}_j \sim G_0, \quad j = 1, 2, \dots, \quad (4d)$$

where $z(t_i)$ denotes the random component index, and each of the gaps $\epsilon_j = \tau_j - \tau_{j-1}$ are i.i.d. from an exponential $\text{Ex}(\lambda)$ prior distribution, with $\tau_0 := 0$. The exponential distribution for creating the intervals is not essential but there seems little reason to make it more complicated.

Interestingly, neither the geometric model nor the NRM model specify a mixing distribution G_t with weights that satisfy the key property, previously described. Figure 1 illustrates this fact. Specifically, for the geometric model, the figure shows samples of the random component index $z(t) \sim \Pr(z(t) = j) \propto \omega_j(t)$, over a convergent sequence of times $t = t_{l-1} + 1/l^2$, for $l = 1, 2, \dots, 1000$, with $t_0 = 0$. These samples are presented for different choices of prior parameters in this model, namely $b = 1, 10, 30, 50$, along with $a = c = 1$. As the figure shows, as t converges to time 1.6439, the random variable $z(t)$ does not converge to a single value. Instead, the random variable displays a degree of uncertainty about the component (kernel) density at that time.

Now, we formally show how our time series model satisfies the property, whereas the geometric model and the NRM models do not. For our model for which we have based on the

$$w_j(t) = \mathbf{1}(t \in A_j)$$

and

$$A_j = (\tau_{j-1}, \tau_j)$$

with

$$\tau_j = \tau_{j-1} + \epsilon_j$$

where the (ϵ_j) are independent and identically distributed exponential random variables with parameter λ , it is straightforward to show that

$$\sum_{j=1}^{\infty} w_j(t) w_j(t+h) = \begin{cases} 1 & \text{with probability } e^{-h} \\ 0 & \text{with probability } 1 - e^{-h}. \end{cases}$$

This follows since we need $t, t+h \in A_j$. Hence, it is seen that

$$\mathbb{E} \left\{ \sum_{j=1}^{\infty} w_j(t) w_j(t+h) \right\} = e^{-h} \rightarrow 1 \quad \text{as } h \rightarrow 0.$$

For the geometric model, we have

$$\mathbb{E} \left\{ \sum_{j=1}^{\infty} w_j(t) w_j(t+h) \right\}$$

given by

$$\mathbb{E} \left\{ \sum_{j=1}^{\infty} \lambda_t (1 - \lambda_t)^{j-1} \lambda_{t+h} (1 - \lambda_{t+h})^{j-1} \right\}$$

which is

$$\mathbb{E} \left\{ \frac{\lambda_t \lambda_{t+h}}{\lambda_t + \lambda_{t+h} - \lambda_t \lambda_{t+h}} \right\}.$$

This is strictly less than one due to the fact that λ_t and λ_{t+h} are less than 1.

Finally, the NRM model also has

$$\mathbb{E} \left\{ \sum_{j=1}^{\infty} w_j(t) w_j(t+h) \right\} < 1,$$

and this result follows from the proof of his Theorem 2, which appears in the Appendix of his paper.

4 Discussion

In summary, we advocate a specific property for mixture models for continuous time series. Namely, that as the time $t+h$ approaches the limit $h \rightarrow 0$, the model should certainly identify a single component index $z(t)$, and hence a single component density $f(y_t | \theta_{z(t)})$ of the dependent response Y_t . In other words, there is no strong reason why one should specify a time-series model that allows the component density to drastically change, as time goes through incrementally smaller changes. In essence we are not asking for G_t to be close to G_{t+h} , though this is given, but a rather weak condition; rather we are asking that θ_t and

θ_{t+h} are close in probability, which approaches 1 as $h \rightarrow 0$.

Interestingly, we have shown that two Bayesian nonparametric (infinite-mixture) models fail this sensible property. In contrast, we have shown that for a mixture model to satisfy the property, it must be of the form given in equation (4). This implies that the mixture model must be a Bayesian multiple change-point model (e.g., Barry & Hartigan, 1993; Chib, 1998), having infinitely-many change-point parameters $\tau_{j-1} < \tau_j$, $j = 1, 2, \dots$. Then, these results may encourage future developments in Bayesian nonparametric models for continuous time series, more in terms of multiple change point modeling.

5 Acknowledgements

This research is supported by National Science Foundation Grant SES-1156372, from the Program in Methodology, Measurement, and Statistics.

Figure 1: For the geometric time series model, the log of samples of component index $z(t) \sim \Pr(z(t) = j) \propto \omega_j(t)$, over a convergent sequence of times $t = t_{l-1} + 1/l^2$, for $l = 1, 2, \dots, 1000$, with $t_0 = 0$. The component index samples are shown for a range of choices of prior parameters, $b = 1, 10, 30, 50$, along with $a = c = 1$.

